Econometric Theory, **31**, 2015, 93–114.
doi:10.1017/S0266466614000218

MODEL DISCOVERY AND TRYGVE HAAVELMO'S LEGACY

DAVID F. HENDRY
University of Oxford

SØREN JOHANSEN
University of Copenhagen
and
Aarhus University

Trygve Haavelmo's *Probability Approach* aimed to implement economic theories, but he later recognized their incompleteness. Although he did not explicitly consider model selection, we apply it when theory-relevant variables, $\{\mathbf{x}_t\}$, are retained without selection while selecting other candidate variables, $\{\mathbf{w}_t\}$. Under the null that the $\{\mathbf{w}_t\}$ are irrelevant, by orthogonalizing with respect to the $\{\mathbf{x}_t\}$, the estimator distributions of the $\mathbf{x}_t$'s parameters are unaffected by selection even for more variables than observations and for endogenous variables. Under the alternative, when the joint model nests the generating process, an improved outcome results from selection. This implements Haavelmo's program relatively costlessly.

1. INTRODUCTION

Toward the end of his career, in conversations with the first author and Eilev Jansen, Trygve Haavelmo confessed himself dissatisfied with the implementation of his 'probability approach in econometrics' research program promulgated in Haavelmo (1944). This was not because of his econometric and statistical proposals, which have withstood the test of time to such an extent that his then novel framework is now taken for granted. Rather, he felt that the available economic theory of the time was not up to the challenge envisaged and had previously stated that view in his presidential address to the Econometric Society, published as Haavelmo (1958). Implementing his view, he had spent much of the intervening period on improving economic reasoning, mainly published in Norwegian: see, e.g., Bjerkholt (2005, 2007) and Moene and Rødseth (1991). To quote Haavelmo (1989):

The first author was supported in part by grants from the Open Society Foundations and the Oxford Martin School, and the second author by the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation). We are grateful to Jennifer L. Castle, David Cox, Jurgen A. Doornik, Neil R. Ericsson, Grayham E. Mizon, and Bent Nielsen for helpful discussions, and to Olav Bjerkholt, Peter C.B. Phillips, and two anonymous referees for useful comments. Address correspondence to David F. Hendry, Institute for New Economic Thinking, University of Oxford, Eagle House, Walton Well Road, Oxford OX2 6ED, UK. email: david.hendry@nuffield.ox.ac.uk

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

> "The basis of econometrics, the economic theories that we had been led to believe in by our forefathers, were perhaps not good enough. It is quite obvious that if the theories we build to simulate actual economic life are not sufficiently realistic, that is, if the data we get to work on in practice are not produced the way that economic theories suggest, then it is rather meaningless to confront actual observations with relations that describe something else."

Cowles Commission research, exemplified by Koopmans (1950) and Hood and Koopmans (1953), was also predicated on the assumption that complete and valid economic theory circumvented issues of data-based model choice by imposing correct restrictions on well-specified models. That viewpoint was embedded in the manifesto by Koopmans (1947), albeit that Koopmans referred to economics as at a 'Newtonian' stage—long after Einstein's relativity theory had been confirmed: see Eddington (1928). In a nutshell, economic analysis formulated the correct model and econometrics merely had the task of estimating its parameters of interest from the best available data, which set the scene for the prevailing orthodoxy still dominant today, where 'data-based' empirical models are readily dismissed as 'measurement without theory' despite the reply by Vining (1949). Indeed, in Pagan (1987), that dominant approach was not one of those criticized. The proposals in both Haavelmo (1944) and Koopmans (1947) share the methodology that economics proposes and evidence disposes, an early implementation of 'conjectures and refutations': see Popper (1963). While certainly an aspect of the scientific enterprise, generating an endless sequence of rejected hypotheses does not seem an efficient route to knowledge acquisition, nor is it characteristic of much of scientific practice. Rather, the earlier 'logic of scientific discovery' in Popper (1959) (a serendipitous mistranslation of Popper's 1935 *Logik der Forschung*) suggests a more productive methodology, namely guiding empirical discovery by the best available theoretical ideas, but always being prepared to have those ideas rejected by sufficiently strong counter evidence.

As a technical body of knowledge, economic analysis has advanced greatly since the 1940s, ironically sowing the very seeds that destroyed empirical findings reliant on imposing earlier, and now defunct, theories on the data: see Hendry (2009). Thus, Haavelmo was correct in the weakness he perceived. However, the counter argument to not imposing a theory-based specification is the manifest inadequacy of short, dependent, aggregated, inaccurate, and heterogeneous time-series data, often subject to extensive revision. Consequently, it seems essential to nest 'theory-driven' and 'data-driven' approaches in order to retain theory insights when exploring empirical interrelations, while also evaluating both the theory and the data evidence: see Hendry (2011a). An approach to doing so is the central theme of this paper.

Econometrics has developed almost out of recognition in the last 75 years or so, with a vast array of new problems tackled. Haavelmo (1944) provided no explicit treatment of 'model selection' in the modern sense of choosing one from

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

a potentially large set of possible representations, nor of alternative approaches to doing so. Indeed, the words 'selection' and 'choice' refer to which theory model to adopt and what functions, variables, and parameters it is to be specified in terms of. 'Testing' in Haavelmo's sense seems to include theory-model selection by checking the 'goodness' of the match of a theory model to the evidence, as well as hypothesis and misspecification testing: for commentaries, see Krüger et al. (1987), Hendry et al. (1989), Spanos (1989), and Juselius (1993).

Rather, Haavelmo (1944) was primarily concerned with what sort of measurement and inference methods follow from the views econometricians hold about the nature of the world represented in their theories and in the data, and what are the right tools (designs of experiments) to uncover economic structures. That in turn led Haavelmo to consider invariance, expressed as "elements of invariance in a relation between real phenomena." In Haavelmo (1944, Ch. 2), he distinguishes explicitly between the constancy of empirical relations and the autonomy of theoretical ones, initiated by Frisch (1938):[1] see later contributions by Aldrich (1989, 1994). Although expectations, a major concern of many economists after Lucas (1976), only occur once in a supply function for savings, he discusses the various levels of autonomy in that theory model—to quote: "we cannot know how far this property of autonomy would in fact be true." Below, we address the issue of superexogeneity: see Engle et al. (1983), Engle and Hendry (1993), and Hendry and Santos (2010), respectively.

The third major contribution of Haavelmo (1944) was formalizing economic forecasting, albeit under the assumption that the model in use was correct and constant, although he highlighted the key role of that last issue.[2] Surprisingly, he did not draw on the prescient work of Smith (1927, 1929). We will only briefly consider forecasting, once the orphan of econometrics, but now a topic of major research, with extensive treatments in four recent handbooks: Clements and Hendry (2002, 2011), Elliott et al. (2006), and Rapach and Wohar (2008).

The unifying theme of this paper is that we are now ready to implement Haavelmo's program, not by imposing a specific subject-matter theory on the data, but via model selection from an initial general specification that tackles the whole gamut of empirical problems jointly, while retaining available theory insights: see Hendry (2011a). We define a theory model to be complete and correct when there are no omitted variables that are relevant in the data generating process (DGP), and no included variables that are irrelevant in the DGP, and all included variables enter with the same functional forms as in the DGP. Almost all theories, including 'general equilibrium,' require a multitude of (usually implicit) *ceteris paribus* statements, especially about the constancy of the theorized relationships, so are unlikely to coincide with the DGP. There are also several important translation issues, as most economic propositions are derived without a unique specification of time units or error terms, whereas statistical relations invariably come at some data frequency and with an error term attached (as expressed by Haavelmo, making "relations more elastic"). Thus, in addition to the deterministic theory postulates, an error $\{\epsilon_t\}$ that is uncorrelated with all the included variables needs to be

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

postulated. Finally, for empirical model selection, $\{\epsilon_t\}$ needs to be independent and identically distributed (IID), perhaps after including features for autocorrelation and heteroskedasticity. We show below that when the theory is complete and correct, such a statistical representation of the theory model can be retained after selection with parameter estimates that have the same distributions as when the model is directly estimated without selection. When the theory is incorrect or incomplete, that will be discovered in a sufficiently large sample when the initial general specification contains some of the functions of the omitted variables; and when the joint model nests the DGP, an improved outcome will result from selection.

The approach can allow for many potentially relevant variables, perhaps at long lags, with possibly nonlinear reactions, facing multiple location shifts and outliers: see Castle et al. (2011) (who provide bibliographic perspective), Johansen and Nielsen (2009), and Castle et al. (2012). The theory model is embedded in the general specification, but is not selected over, whereas other aspects are selected at tight significance levels, so are only retained when highly significant. The general model often has more candidate variables N than observations T, so block contracting and expanding searches are required, as implemented in *Autometrics*: see Doornik (2007, 2009). Once a feasibly estimable model is obtained with $n << T$, it is evaluated for congruence by a range of misspecification tests, and if not rejected, reduction continues till no insignificant variables remain, then that terminal model is checked by encompassing: see Doornik (2008).

The structure of the paper is as follows. Section 2 briefly describes Trygve Haavelmo's legacy. Section 3 discusses the problems of nonstationarity in economic time series where breaks and stochastic trends are ubiquitous. Section 4 considers empirical model discovery in that context, retaining theory insights while allowing a major role for evidence in modeling time series, as theory often abstracts from important sources of data variation due to breaks and stochastic trends. Section 4.1 shows that the distributions of the estimated coefficients of m exogenous variables $\mathbf{x}_t$ are unaffected by model selection when the k irrelevant variables $\mathbf{w}_t$ are orthogonalized with respect to $\mathbf{x}_t$, for $(k+m) << T$, so the general model is estimable. Section 4.3 establishes that the same results apply even when $(k+m) > T$, provided $m << T$. Section 4.4 extends the analysis to a valid theory with endogenous variables and Section 4.5 notes how to assess the validity of the instrumental variables. Section 5 provides a simulation illustration. Section 6 relates the selected model to the problems of economic forecasting; and Section 7 concludes.

2. TRYGVE HAAVELMO'S LEGACY

What is Haavelmo's legacy? Hendry and Morgan (1995) devote a major section of their book to his many contributions, but here, the key aspect was his stressing the need for, and providing the ability to, formalize the joint probability distribution of the data, now called the Haavelmo distribution following Spanos (1989).

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

Morgan (1990, Ch. 8) presents evidence that most pre-1940 econometricians did not believe that probability could be applied to economic data because such data lacked independence (nonrandom sampling) and homogeneity (drawn from changing distributions), both seen as essential prerequisites for a viable statistical analysis. Keynes was among the most vociferous critics of the failure of such assumptions in his review of Tinbergen (1939): see Keynes (1939, 1940) and the retort in Marschak and Lange (1940).

However, neither independence nor homogeneity is in fact required of data to develop a viable statistical analysis, as formalized by Doob (1953). Let the set of stochastic variables to be analyzed be denoted $\mathbf{X}_T^1 = (\mathbf{x}_1 \dots \mathbf{x}_T)$, then their joint data density $\mathsf{f}_{\mathsf{X}}(\cdot)$ conditional on the past and for parameters $\mathbf{\Lambda}_T^1 = (\boldsymbol{\lambda}_1 \dots \boldsymbol{\lambda}_T)$ can be sequentially factorized as:

$$\mathsf{f}_{\mathsf{X}}\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \mathbf{Q}_T^1, \mathbf{\Lambda}_T^1\right) = \prod_{t=1}^{T} \mathsf{f}_{\mathsf{x}_t}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}^1, \mathbf{X}_0, \mathbf{q}_t, \boldsymbol{\lambda}_t\right), \tag{1}$$

where $\mathbf{Q}_T^1 = (\mathbf{q}_1 \dots \mathbf{q}_T)$ are deterministic terms, and $\mathbf{X}_0 = (\dots \mathbf{x}_0)$. Let:

$$\boldsymbol{\epsilon}_t = \mathbf{x}_t - E_{\mathsf{f}_{\mathsf{x}_t}}\left[\mathbf{x}_t \mid \mathbf{X}_{t-1}^1, \mathbf{X}_0, \mathbf{q}_t, \boldsymbol{\lambda}_t\right],$$

then $E_{\mathsf{f}_{\mathsf{x}_t}}\left[\boldsymbol{\epsilon}_t | \mathbf{X}_{t-1}^1, \mathbf{X}_0, \mathbf{q}_t, \boldsymbol{\lambda}_t\right] = \mathbf{0}$, so $\{\boldsymbol{\epsilon}_t\}$ is a (possibly heteroskedastic) mean innovation process (MIP) against the appropriate filter, and a martingale difference sequence as $E_{\mathsf{f}_{\mathsf{x}_t}}\left[\boldsymbol{\epsilon}_t | \mathbf{E}_{t-1}\right] = \mathbf{0}$ for $\mathbf{E}_{t-1} = (\dots \boldsymbol{\epsilon}_1 \dots \boldsymbol{\epsilon}_{t-1})$. Since:

$$\log \mathsf{f}_{\mathsf{X}}\left(\mathbf{X}_T^1 \mid \mathbf{X}_0, \mathbf{Q}_T^1, \mathbf{\Lambda}_T^1\right) = \sum_{t=1}^{T} \log \mathsf{f}_{\mathsf{x}_t}\left(\mathbf{x}_t \mid \mathbf{X}_{t-1}^1, \mathbf{X}_0, \mathbf{q}_t, \boldsymbol{\lambda}_t\right), \tag{2}$$

the log sequential densities in (2) provide a basis for laws of large numbers and central limit theorems.

Thus, the model specification problem is to characterize the sequential densities $\mathsf{f}_{\mathsf{x}_t}(\mathbf{x}_t | \cdot)$, as these are not estimable. The set of variables in $\{\mathbf{x}_t\}$ is usually postulated by prior theoretical analysis, and there will also be theories of how their elements are connected in $\{\boldsymbol{\lambda}_t\}$. But absent omniscience, many difficult modeling issues remain: precisely what are the formulations of $\mathsf{f}_{\mathsf{x}_t}(\mathbf{x}_t | \cdot)$; what are the (perhaps evolving) connections between those $\mathbf{x}_t$; what is the shortest lag length, s, such that $E_{\mathsf{f}_{\mathsf{x}_t}}\left[\boldsymbol{\epsilon}_t | \mathbf{X}_{t-1}^{t-s}, \mathbf{X}_0, \mathbf{q}_t, \boldsymbol{\lambda}_t\right] = \mathbf{0}$ so $\{\boldsymbol{\epsilon}_t\}$ remains a MIP; what are the detailed (possibly nonlinear) equation specifications; what are the underlying constancies and invariances of the sequential distributions; and what is the composition of $\{\mathbf{q}_t\}$? Finally, going outside the postulated framework, we consider whether the $\mathbf{x}_t$ comprise all the relevant variables, so $\mathsf{f}_{\mathsf{X}}(\cdot)$ is the DGP or is just the local DGP (LDGP: the derived DGP in the space of the variables under analysis: see Hendry, 2009).

Haavelmo (1944) essentially formalized this approach for constant $\mathsf{f}_{\mathsf{x}}(\mathbf{x}_t | \cdot)$, but still allowed for nonconstancy through changing $\boldsymbol{\lambda}$ or by location-shift dummies

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

in $\mathbf{Q}_T^1$. Indeed, estimation and inference only become viable once a model captures changes in $\mathsf{f}_{\mathsf{X}_t}(\mathbf{x}_t|\cdot)$ by such mean shifts, or models of its moments. Two important forms of nonstationarity have since received considerable attention, so we review those next.

3. THE PROBLEMS OF NONSTATIONARITY

Stochastic trends, breaks, and regime shifts are ubiquitous in economic time series and were already widely recognized by the time of Haavelmo (1944). From at least Hooker (1901), through Yule (1926), integrated time series were known to be problematic when analyzed by conventional statistical methods such as static regressions between levels. However, like most of his contemporaries, Haavelmo seems to have been unaware of the implications of the pioneering work of Smith (1926), recently brought to light by Mills (2010). Smith (1926) effectively solved the problem of nonsense regressions for I(1) time series by nesting levels and differences in what would now be recognized as an unrestricted equilibrium-correction model (EqCM), although he in turn does not cite Yule (1926), so seems to have been unaware of the distributional implications of I(1) data. Since then, integrated data have received considerable treatment via unit roots and cointegration: see among many others, Dickey and Fuller (1979), Granger (1981), Phillips (1986, 1991), Engle and Granger (1987), and Johansen (1988, 1995). Nevertheless, such developments are extensions of the Haavelmo framework, rather than replacements, even though the resulting statistical models and methods were not envisaged at the time. Sims et al. (1990) show that although t- and F-statistics have nonstandard distributions even in correctly specified models fitted to I(1) data, conventional critical values are valid for most tests provided the model can be rewritten in terms of stationary variables: see Toda and Phillips (1993). The test for a reduction to I(0) always needs 'unit-root' critical values. Thus, model selection can be applied to most conditional specifications on the original data, as a test of a unit-root hypothesis in a level specification will almost never be insignificant. The one potentially problematic setting is when several lags of an irrelevant unmodeled I(1) variable are included: when all but one of the lags have been eliminated during selection, a t-test on the last one cannot be written in the way Sims et al. (1990) propose. Thus, slightly tighter critical values seem advisable if that is suspected. Omtzig (2002), Kurcewicz and Mycielski (2003), and Liao and Phillips (2012) have analyzed algorithms for automatic selection of cointegration vectors, but we do not address that issue further here.

The other main nonstationarity comprises changes in distributions, including location shifts (changed means), breaks in unconditional variances and other moments; shape changes; shifts in parameters of relationships between variables, etc. Clements and Hendry (1998) and many other papers show that the most pernicious changes for forecasts are location shifts, which lead to systematic forecast failure for all members of the huge class of EqCMs, as such models correct back to the previous location whatever the new mean. Correspondingly, location shifts

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

are relatively easy to detect, whereas many other (mean-zero) changes can be difficult to detect till long after their occurrence. Thus, we will focus on handling location shifts in the context of model selection by including in the candidate variables a saturating set of impulse indicators $1_{\{j=t\}}$ $t = 1, \ldots, T$ (denoted IIS for impulse-indicator saturation). Johansen and Nielsen (2009) show that the split-half approach in Hendry et al. (2008) can be generalized to dynamic regression models, and despite including T additional candidate indicators, the limiting distributions of the retained parameters of interest have the same characteristics as if IIS had not been undertaken, other than a small increase in the limiting variance dependent on the critical value used for selection.

When there are relevant variables, $\mathbf{w}_t$, excluded from the conditional DGP of a variable of interest, y_t, but correlated with included variables, $\mathbf{x}_t$, a mean shift in the included variables' DGP alters intercepts in y_t's LDGP. To see this, let the conditional DGP of y_t be:

$$y_t = \boldsymbol{\beta}' \mathbf{x}_t + \boldsymbol{\gamma}' \mathbf{w}_t + \epsilon_t, \tag{3}$$

where $\epsilon_t \sim \mathsf{IID}[0, \sigma_\epsilon^2]$, $\boldsymbol{\beta} \neq \mathbf{0}$, $\boldsymbol{\gamma} \neq \mathbf{0}$, and ϵ_t is independent of the strongly exogenous variables $\mathbf{x}_t$ and $\mathbf{w}_t$ which are linked in-sample by the projection:

$$\mathbf{w}_t = \boldsymbol{\psi} + \boldsymbol{\Psi} \mathbf{x}_t + \mathbf{v}_t \tag{4}$$

when $E\left[\mathbf{x}_t \mathbf{v}_t'\right] = \mathbf{0}$, and $\boldsymbol{\Psi} \neq \mathbf{0}$. Let the variables included in the conditional model $y_t | \mathbf{x}_t$ have $E[\mathbf{x}_t] = \boldsymbol{\delta}_1$ in-sample. When the relevant variables $\mathbf{w}_t$ are excluded, the conditional LDGP of y_t given $\mathbf{x}_t$ derived from (3) and (4) becomes:

$$y_t = \boldsymbol{\gamma}' \boldsymbol{\psi} + \left(\boldsymbol{\beta}' + \boldsymbol{\gamma}' \boldsymbol{\Psi}\right) \boldsymbol{\delta}_1 + \left(\boldsymbol{\beta}' + \boldsymbol{\gamma}' \boldsymbol{\Psi}\right) (\mathbf{x}_t - \boldsymbol{\delta}_1) + \boldsymbol{\gamma}' \mathbf{v}_t + \epsilon_t, \tag{5}$$

where we have separated out the means so that $E[\mathbf{x}_t - \boldsymbol{\delta}_1] = \mathbf{0}$ and hence $E[y_t] = \boldsymbol{\gamma}' \boldsymbol{\psi} + \left(\boldsymbol{\beta}' + \boldsymbol{\gamma}' \boldsymbol{\Psi}\right) \boldsymbol{\delta}_1$. The model corresponding to a regression of y_t on $(1, \mathbf{x}_t)$:

$$y_t = \lambda_0 + \boldsymbol{\lambda}_1' \mathbf{x}_t + e_t \tag{6}$$

matches the conditional LDGP (5) with $\lambda_0 = \boldsymbol{\gamma}' \boldsymbol{\psi}$ and $\boldsymbol{\lambda}_1 = \boldsymbol{\beta} + \boldsymbol{\Psi}' \boldsymbol{\gamma}$, so estimation thereof will deliver estimates of the parameters in (5).

When some $\mathbf{x}_t$ are policy variables, changes in their means at time $T+1$ would alter $E[\mathbf{x}_T]$ from $\boldsymbol{\delta}_1$ to $E\left[\mathbf{x}_{T+1}\right] = \boldsymbol{\delta}_2$. The actual outcome from (3), even if $E[\mathbf{w}]$ does not also change—so (4) shifts—would be an average change in y of:

$$E\left[y_{T+1}\right] - E[y_T] = \boldsymbol{\beta}' (\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1). \tag{7}$$

Using E_{M} to denote an expectation based on (6), the shift in $\mathbf{x}$ would produce an average anticipated change of:

$$E_{\mathsf{M}}\left[y_{T+1}\right] - E_{\mathsf{M}}[y_T] = \boldsymbol{\lambda}_1' (\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1) = \left(\boldsymbol{\beta}' + \boldsymbol{\gamma}' \boldsymbol{\Psi}\right) (\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1) \tag{8}$$

resulting in an unexpected location shift of $\boldsymbol{\gamma}' \boldsymbol{\Psi} (\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1)$, augmented by any concomitant changes in $E\left[\mathbf{w}_{T+1}\right]$. This could lead to an adverse effect of the policy,

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

induced in (5) by the shift in $\mathbf{x}_T$ interacting with the misspecification of omitting a relevant correlated variable, so is a violation of the superexogeneity of $\mathbf{x}_t$: see Engle and Hendry (1993). The solution of including $\{\mathbf{w}_t\}$ is addressed in the next section. Here we note that similar shifts may have already occurred in-sample, so before policy changes are implemented, a prior test for the superexogeneity of $\mathbf{x}_t$ is imperative to reveal such an impending problem. The test for superexogeneity in Hendry and Santos (2010) uses IIS in an automatically created vector autoregressive model of $\{\mathbf{x}_t\}$ to ascertain the occurrence and timing of location shifts like $(\boldsymbol{\delta}_2 - \boldsymbol{\delta}_1)$ in-sample, then tests whether those enter the conditional model $y_t|\mathbf{x}_t$, where their significance would reject the hypothesis of superexogeneity. Such co-breaking failures can help determine the sources and likely consequences of shifts before implementing policies.

4. EMPIRICAL MODEL DISCOVERY

Model selection, aka "Data Mining," has not been well regarded: see, among others, the criticisms in the discussion of Coen et al. (1969), or the critical analyses in Bock et al. (1973), Leamer (1978), Lovell (1983), Chatfield (1995), and Leeb and Pötscher (2005), who highlight excess retention of irrelevant variables and poor location of the DGP, with biased estimates after selection. However, there are counter discussions in Sargan (2001), Hoover and Perez (2000), and Hendry (2000) who accept that many methods of selection do not work well, especially 1-step expanding searches such as stepwise, but it is a *non sequitur* to conclude that no methods can work well: Anderson (1962) demonstrated the benefits of contracting over expanding searches. Moreover, most analytical and simulation results are for very different settings than the one that confronts empirical economics, namely assuming that only theory-based variables are relevant and provide a comprehensive specification, but nevertheless selection is undertaken. In practice, some of the key features driving data variability are absent from any theory and require to be discovered empirically. Recent advances follow from the first generation of automatic selection methods in Hoover and Perez (1999), then Hendry and Krolzig (2005), leading to the third generation implemented in *Autometrics* as noted above.

There are many ways to judge the performance of such algorithms, but some basic requirements are that the null retention rate of irrelevant variables (gauge) is close to the nominal size, α, set for selection tests; the retention rate of relevant variables (potency) is not far below the corresponding power of the equivalent test in the relevant LDGP; and that the LDGP is located almost as often starting from the postulated initial general model as when starting from the LDGP itself: see Castle et al. (2011). Gauge and potency differ from size and power, respectively, not least because algorithms can retain variables whose estimated coefficients are insignificant on the selection statistic. Recent algorithms handle more variables, N, than observations, T, to tackle general nonlinear approximations and multiple breaks using IIS. Once $N > T$, model selection is both essential and unavoidable

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

without incredible claims to perfect *a priori* knowledge of the world. Analytical comparisons with other procedures are difficult, although operating characteristics such as gauge, potency, and relative performance at locating the LDGP can be evaluated by simulation.

Economic theories are often fitted directly to data to avoid possible 'model-selection biases.' This is an excellent strategy when the theory is complete and correct, but less successful otherwise. We show that by embedding a theory model that specifies the correct set of m relevant exogenous variables, $\mathbf{x}_t$, within a larger set of $m+k$ candidate variables, $(\mathbf{x}_t, \mathbf{w}_t)$, our approach to selecting over the second set by their statistical significance can be undertaken without affecting the theory parameters' estimator distributions. This strategy keeps the same theory-parameter estimates as direct fitting when the theory is correct, yet protects against the theory being under-specified when some $\mathbf{w}_t$ are relevant.

4.1. Selection When Retaining a Valid Theory

Consider a theory model which correctly matches the data-generating process (DGP) by specifying that:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \epsilon_t, \tag{9}$$

where $\epsilon_t \sim \mathsf{IID}[0, \sigma_\epsilon^2]$ over $t = 1, \ldots, T$, and ϵ_t is independent of the m strongly exogenous variables $\{\mathbf{x}_1, \ldots, \mathbf{x}_t\}$, assumed to satisfy:

$$T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' \xrightarrow{P} \boldsymbol{\Sigma}_{xx},$$

which is positive definite, and:

$$T^{1/2}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \left(T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1} T^{-1/2}\sum_{t=1}^{T}\mathbf{x}_t\epsilon_t \xrightarrow{D} \mathsf{N}_m\left[\mathbf{0}, \sigma_\epsilon^2\boldsymbol{\Sigma}_{xx}^{-1}\right], \tag{10}$$

where $\boldsymbol{\beta}_0$ is the constant population parameter.

However, an investigator may be willing to contemplate the possibility that an additional set of k exogenous variables $\mathbf{w}_t$ also influences y_t, so postulates the more general model:

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{w}_t + \epsilon_t, \tag{11}$$

although in fact $\boldsymbol{\gamma}_0 = \mathbf{0}$. The $\mathbf{w}_t$ can be variables known to be exogenous, functions of those, lagged variables in time series, and indicators for outliers or breaks, and we assume the same assumptions as above for $\{\epsilon_t, \mathbf{x}_t, \mathbf{w}_t\}$. The investigator regards the theory in (9) as correct and complete, so wishes to ensure that the $\mathbf{x}_t$

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

are always retained and not selected over (called forced). The issue we address is the possible additional cost of searching over the candidate variables $\mathbf{w}_t$ in (11) when retaining the $\mathbf{x}_t$, rather than directly estimating (9) when $(k+m) << T$.

As a generalization of Frisch and Waugh (1933) (see Davidson and MacKinnon (2004)), $\mathbf{x}_t$ and $\mathbf{w}_t$ can be orthogonalized by first computing:

$$\widehat{\mathbf{\Gamma}} = \left(\sum_{t=1}^{T} \mathbf{w}_t \mathbf{x}_t'\right)\left(\sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t'\right)^{-1} \tag{12}$$

and defining the residuals $\widehat{\mathbf{u}}_t$ by:

$$\mathbf{w}_t = \widehat{\mathbf{\Gamma}}\mathbf{x}_t + \widehat{\mathbf{u}}_t \tag{13}$$

so that:

$$\sum_{t=1}^{T} \mathbf{x}_t \widehat{\mathbf{u}}_t' = \mathbf{0}. \tag{14}$$

Using (13) in (11):

$$\begin{aligned} y_t &= \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\gamma}'\mathbf{w}_t + \epsilon_t = \boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\gamma}'\left(\widehat{\mathbf{\Gamma}}\mathbf{x}_t + \widehat{\mathbf{u}}_t\right) + \epsilon_t \\ &= \boldsymbol{\beta}_+'\mathbf{x}_t + \boldsymbol{\gamma}'\widehat{\mathbf{u}}_t + \epsilon_t, \end{aligned} \tag{15}$$

where $\boldsymbol{\beta}_+ = \boldsymbol{\beta} + \widehat{\mathbf{\Gamma}}'\boldsymbol{\gamma}$. Note that $\boldsymbol{\beta}_{0+} = \boldsymbol{\beta}_0$ because $\boldsymbol{\gamma}_0 = \mathbf{0}$. Consequently, as (9) is the DGP, by orthogonality from (14):

$$\begin{aligned} T^{1/2}\begin{pmatrix} \widetilde{\boldsymbol{\beta}}_+ - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\gamma}} \end{pmatrix} &= \begin{pmatrix} T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' & T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\widehat{\mathbf{u}}_t' \\ T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\mathbf{x}_t' & T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_t' \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2}\sum_{t=1}^{T}\mathbf{x}_t\epsilon_t \\ T^{-1/2}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\epsilon_t \end{pmatrix} \\ &= \begin{pmatrix} \left(T^{-1}\sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t'\right)^{-1} T^{-1/2}\sum_{t=1}^{T}\mathbf{x}_t\epsilon_t \\ \left(T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}_t'\right)^{-1} T^{-1/2}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\epsilon_t \end{pmatrix} \\ &\xrightarrow{D} \mathsf{N}_{m+k}\left[\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \sigma_\epsilon^2\begin{pmatrix} \mathbf{\Sigma}_{xx}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{ww|x}^{-1}\end{pmatrix}\right]. \end{aligned} \tag{16}$$

Thus, the estimator $\widetilde{\boldsymbol{\beta}}_+$ in (16) is identical to $\widehat{\boldsymbol{\beta}}$ in (10), independently of the inclusion or exclusion of any or all of the $\widehat{\mathbf{u}}_t$. Even after selection over the $\widehat{\mathbf{u}}_t$ at significance level α, and corresponding critical value c_α, say, by sequential t-tests on each $\widetilde{\gamma}_i$, the theory-parameter estimator is unaffected by retaining significant $\widehat{\mathbf{u}}_t$ when $\boldsymbol{\gamma}_0 = \mathbf{0}$. For a Gaussian distribution and fixed regressors, the estimator $\widetilde{\boldsymbol{\beta}}_+ = \widehat{\boldsymbol{\beta}}$ is statistically independent of the test statistics used to select, and therefore the distribution of $\widehat{\boldsymbol{\beta}}$ is unaffected by the selection over $\mathbf{w}_t$ when in fact they are irrelevant.

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

The main cost of selection is the chance retention of some $\widehat{u}_{i,t}$, which may mislead on the validity of the theory model. If all $\widehat{u}_{i,t}$ are irrelevant, then on average αk will be retained by chance, with estimated coefficient $\widetilde{\gamma}_i$, where:

$$\left|\mathsf{t}_{\gamma_i=0}\right| = \frac{|\widetilde{\gamma}_i|}{\mathsf{SE}[\widetilde{\gamma}_i]} \geq c_\alpha. \tag{17}$$

Setting $\alpha = \min[1/k, 1/T, 1\%]$ is an appealing rule. When $T = 100$ and $k = T/4 = 25$, say, then because $k\alpha = 0.25$, the probability of retaining more than one irrelevant variable is:

$$\mathsf{p}_1 = 1 - \sum_{i=0}^{1} \frac{(0.25)^i}{i!} e^{-0.25} \simeq 2.6\%. \tag{18}$$

Moreover, under normality, for $h > 2/c_\alpha$ then:

$$P\left(\left|\mathsf{t}_{\gamma_i=0}\right| \geq h c_\alpha \mid \mathsf{H}_0\right) \leq \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{h^2}{2} c_\alpha^2\right), \tag{19}$$

which is 0.01% at $h = 1.5$ and $c_{0.01} = 2.65$. Thus, it is unlikely any $\left|\mathsf{t}_{\gamma_i=0}\right|$ will be larger than $1.5c_\alpha$. Thus, only reject a theory when more than one of the $\widehat{u}_{i,t}$ are retained, or when one is more significant than $1.5c_\alpha$. In empirical applications, a simpler decision rule for rejecting the theory is that an (approximate) F-test on all the selected unforced variables is significant at α_2, where $\alpha_2 < \alpha_1$.

Also, under the null that $\boldsymbol{\gamma}_0 = \mathbf{0}$, an unbiased estimated error variance is:

$$\widetilde{\sigma}_\epsilon^2 = (T-m)^{-1} \sum_{t=1}^{T} \left(y_t - \widetilde{\boldsymbol{\beta}}'_+ \mathbf{x}_t\right)^2 \tag{20}$$

so that the *estimated* distribution of $\widetilde{\boldsymbol{\beta}}_+$ in (16) is correctly scaled, although under the alternative, (20) will be an overestimate. Estimates of γ_i can be approximately bias corrected after their chance retention, as in Hendry and Krolzig (2005), which markedly reduces their mean square errors.

Against these possible costs, the main benefits are that the theory model is tested simultaneously against the wide range of alternatives in $\mathbf{w}_t$; and when the theory is incomplete, the selected model will be less misspecified relative to direct estimation, as we now discuss.

4.2. Retaining an Incomplete or Invalid Theory

Under the alternative that $\boldsymbol{\gamma}_0 \neq \mathbf{0}$, directly fitting (9) will result in estimating the coefficient $\boldsymbol{\beta}_0 + \widehat{\boldsymbol{\Gamma}}' \boldsymbol{\gamma}_0$. Even when (11) nests the DGP, from (15), $\boldsymbol{\beta}_0 + \widehat{\boldsymbol{\Gamma}}' \boldsymbol{\gamma}_0$ will also be the coefficient of $\mathbf{x}_t$. However, when (11) does nest the DGP, selection can improve the final model relative to (9), as in Castle et al. (2011).

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

While retaining $\mathbf{x}_t$ when selecting from (15) will then deliver an incorrect estimate of $\boldsymbol{\beta}_0$, relevant $\widehat{u}_{i,t}$ with large noncentralities will usually be retained, this time correctly, rejecting the validity of the theory model and providing a better approximation to the LDGP. With small noncentralities, no $\widehat{u}_{i,t}$ may be retained even when $\gamma_0 \neq \mathbf{0}$, so an incorrect selection results, matching the direct fit of the theory model. If desired, an estimate of $\boldsymbol{\beta}_0$ could be derived from the estimated coefficient of $\mathbf{x}_t$ in (15), namely:

$$\widetilde{\boldsymbol{\beta}}_+ = \widetilde{\left(\boldsymbol{\beta}_0 + \widehat{\boldsymbol{\Gamma}}' \boldsymbol{\gamma}_0\right)}$$

using $\widetilde{\boldsymbol{\beta}}_0 = \widetilde{\boldsymbol{\beta}}_+ - \widetilde{\left(\widehat{\boldsymbol{\Gamma}}' \boldsymbol{\gamma}_0\right)}$ since $\widetilde{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\Gamma}}$ are known.

4.3. More Candidate Variables than Observations

The analytic approach in Johansen and Nielsen (2009) to understanding IIS also applies under the null for $k = T$ IID mutually orthogonal candidate regressors. Add the first $k/2$ of the variables and select at significance level $\alpha = 1/T = 1/k$. Record which are significant, then drop them. Now add the second block of $k/2$, again selecting at significance level $\alpha = 1/k$, and record which are significant in that subset. Finally, combine the recorded variables from the two stages (if any), and select again at significance level $\alpha = 1/k$. At both substeps, on average $\alpha k/2 = 1/2$ a variable will be retained by chance, so on average $\alpha k = 1$ will be retained from the combined stage. Under the null, one degree of freedom is lost on average. Instead of just the split-half approach, a combination of expanding and contracting multiple block searches is implemented in (e.g.) *Autometrics* as described in Doornik (2009) and Doornik and Hendry (2009).

In the more important setting where the model also has theory-based relevant variables to be retained, so $k + m = N > T$, orthogonalize the relevant variables with respect to the other candidates as above, but in blocks. Under the null, doing so has no impact on the parameters of the retained variables, or their estimates. When $N > T$, divide the N variables into subblocks of smaller than $T/4$ (say), setting $\alpha = 1/N$ overall. The selected model retains the desired subset of m theory-based variables at every stage, and only selects over the putative irrelevant variables at a stringent significance level. The earlier criteria for rejecting the theory model can still be applied.

4.4. Retaining a Valid Theory with Endogenous Variables

When some of the right-hand side variables are potentially endogenous, the theory model is still:

$$y_t = \boldsymbol{\beta}' \mathbf{x}_t + \epsilon_t, \tag{21}$$

where $\mathbf{x}_t$ is $m \times 1$, and $\epsilon_t \sim \mathsf{IID}[0, \sigma_\epsilon^2]$, but now ϵ_t is independent of the $n \geq m$ instrumental variables $\mathbf{z}_1, \ldots, \mathbf{z}_t$ where $(m+n) < T$. The partial DGP for the

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

variables $(y_t, \mathbf{x}_t)$ given $\mathbf{z}_t$ has the form:

$$y_t = \boldsymbol{\beta}'\boldsymbol{\Pi}\mathbf{z}_t + \eta_t$$
$$\mathbf{x}_t = \boldsymbol{\Pi}\mathbf{z}_t + \boldsymbol{\xi}_t,$$

where $(\eta_t, \boldsymbol{\xi}_t)$ are $\mathsf{IID}[\mathbf{0}, \boldsymbol{\Omega}]$ with:

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_\eta^2 & \boldsymbol{\sigma}'_{\eta\xi} \\ \boldsymbol{\sigma}_{\xi\eta} & \boldsymbol{\Omega}_\xi \end{pmatrix}$$

and $(\eta_t, \boldsymbol{\xi}_t)$ is independent of $\mathbf{z}_1, \ldots, \mathbf{z}_t$, but $\epsilon_t = y_t - \boldsymbol{\beta}'\mathbf{x}_t = \eta_t - \boldsymbol{\beta}'\boldsymbol{\xi}_t$ is correlated with $\mathbf{x}_t$ as $\mathsf{Cov}[\mathbf{x}_t\epsilon_t] = \boldsymbol{\sigma}_{\xi\eta} - \boldsymbol{\Omega}_\xi\boldsymbol{\beta}$.

Then instrumental variables estimation of (21) coincides with two-stage least squares (2SLS) and delivers:

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \left[\left(\sum_{t=1}^T \mathbf{x}_t\mathbf{z}'_t\right)\left(\sum_{t=1}^T \mathbf{z}_t\mathbf{z}'_t\right)^{-1}\left(\sum_{t=1}^T \mathbf{z}_t\mathbf{x}'_t\right)\right]^{-1} \times \left(\sum_{t=1}^T \mathbf{x}_t\mathbf{z}'_t\right)\left(\sum_{t=1}^T \mathbf{z}_t\mathbf{z}'_t\right)^{-1}\sum_{t=1}^T \mathbf{z}_t\epsilon_t \tag{22}$$

so that:

$$T^{1/2}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \xrightarrow{D} \mathsf{N}_m\left[\mathbf{0}, \sigma_\epsilon^2\mathbf{Q}^{-1}\right], \tag{23}$$

where we assume:

$$\mathbf{Q} = \operatorname*{plim}_{T\to\infty}\left[\left(\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t\mathbf{z}'_t\right)\left(\frac{1}{T}\sum_{t=1}^T \mathbf{z}_t\mathbf{z}'_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T \mathbf{z}_t\mathbf{x}'_t\right)\right]$$

is positive definite. Let:

$$\widehat{\boldsymbol{\Pi}} = \left(\frac{1}{T}\sum_{t=1}^T \mathbf{x}_t\mathbf{z}'_t\right)\left(\frac{1}{T}\sum_{t=1}^T \mathbf{z}_t\mathbf{z}'_t\right)^{-1}$$

and define:

$$\widehat{\mathbf{x}}_t = \widehat{\boldsymbol{\Pi}}\mathbf{z}_t \text{ with } \widehat{\boldsymbol{\xi}}_t = \mathbf{x}_t - \widehat{\mathbf{x}}_t = \left(\boldsymbol{\Pi} - \widehat{\boldsymbol{\Pi}}\right)\mathbf{z}_t + \boldsymbol{\xi}_t,$$

then a 2SLS reformulation that is algebraically convenient is:

$$y_t = \boldsymbol{\beta}'\widehat{\mathbf{x}}_t + e_t, \tag{24}$$

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

where:

$$e_t = \epsilon_t + \boldsymbol{\beta}'\widehat{\boldsymbol{\xi}}_t = \eta_t + \boldsymbol{\beta}'(\boldsymbol{\xi}_t - \widehat{\boldsymbol{\xi}}_t)$$

so that:

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \widehat{\mathbf{x}}_t e_t = \operatorname*{plim}_{T\to\infty} \widehat{\boldsymbol{\Pi}} \frac{1}{T}\sum_{t=1}^{T} \mathbf{z}_t \left(\eta_t + \boldsymbol{\beta}'\left(\boldsymbol{\xi}_t - \widehat{\boldsymbol{\xi}}_t\right)\right) = \mathbf{0}.$$

When an investigator includes an additional set of k candidate exogenous variables $\mathbf{w}_t$, consider the partial DGP:

$$\begin{aligned} y_t &= \boldsymbol{\beta}'\boldsymbol{\Pi}\mathbf{z}_t + \boldsymbol{\gamma}'\mathbf{w}_t + \eta_t \\ \mathbf{x}_t &= \boldsymbol{\Pi}\mathbf{z}_t + \boldsymbol{\xi}_t, \end{aligned} \tag{25}$$

where $\boldsymbol{\gamma}_0 = \mathbf{0}$, and the $\mathbf{x}_t$ are retained. Since $\boldsymbol{\gamma}_0 = \mathbf{0}$, when the $\widehat{\mathbf{x}}_t = \widehat{\boldsymbol{\Pi}}\mathbf{z}_t$ and $\mathbf{w}_t$ are orthogonalized as in (13), from (25):

$$\begin{aligned} y_t &= \boldsymbol{\beta}'\widehat{\mathbf{x}}_t + \boldsymbol{\gamma}'\mathbf{w}_t + \eta_t + \boldsymbol{\beta}'\left(\boldsymbol{\xi}_t - \widehat{\boldsymbol{\xi}}_t\right) \\ &= \boldsymbol{\beta}'\widehat{\mathbf{x}}_t + \boldsymbol{\gamma}'\left(\widehat{\boldsymbol{\Gamma}}\widehat{\mathbf{x}}_t + \widehat{\mathbf{u}}_t\right) + e_t = \boldsymbol{\beta}'_+\widehat{\mathbf{x}}_t + \boldsymbol{\gamma}'\widehat{\mathbf{u}}_t + e_t. \end{aligned} \tag{26}$$

When (21) is the DGP, by orthogonality from (13):

$$\begin{aligned} T^{1/2}\begin{pmatrix} \widetilde{\boldsymbol{\beta}}_+ - \boldsymbol{\beta}_0 \\ \widetilde{\boldsymbol{\gamma}} \end{pmatrix} &= \begin{pmatrix} T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{x}}_t\widehat{\mathbf{x}}'_t & T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{x}}_t\widehat{\mathbf{u}}'_t \\ T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{x}}'_t & T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}'_t \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2}\sum_{t=1}^{T}\widehat{\mathbf{x}}_t e_t \\ T^{-1/2}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t e_t \end{pmatrix} \\ &= \begin{pmatrix} \left(T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{x}}_t\widehat{\mathbf{x}}'_t\right)^{-1} T^{-1/2}\sum_{t=1}^{T}\widehat{\mathbf{x}}_t e_t \\ \left(T^{-1}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t\widehat{\mathbf{u}}'_t\right)^{-1} T^{-1/2}\sum_{t=1}^{T}\widehat{\mathbf{u}}_t e_t \end{pmatrix} \\ &\xrightarrow{D} \mathsf{N}_{m+k}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \sigma^2_\eta \begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{\widehat{x}\widehat{x}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}^{-1}_{ww|z} \end{pmatrix}\right]. \end{aligned} \tag{27}$$

Thus, the estimator $\widetilde{\boldsymbol{\beta}}_+$ in (27) is again identical to the estimator $\widehat{\boldsymbol{\beta}}$ in (23), independently of the inclusion or exclusion of any or all of the $\widehat{\mathbf{u}}_t$, so is also unaffected by selecting significant $\widehat{u}_{i,t}$.

While model selection for OLS and 2SLS, even with IIS, are simple cases, the idea of embedding a retained theory model in a set of orthogonalized variables is general and could be extended to a wide range of settings, including systems. Under the null, there will be no impact on the estimated parameter distributions relative to direct estimation, and under the alternative that the extended model nests the DGP, improved results will be obtained.

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

4.5. Assessing the Validity of the Instrumental Variables

When the equation is over-identified, the validity of the instrumental variables and any additional candidate regressors can be checked by the usual Durbin–Wu–Hausman test against the most reliable instruments as the baseline: see Durbin (1954), Wu (1973), and Hausman (1978). Alternatively, the least reliable instruments can be added to the theory-based equation: see Hendry (2011b).

5. A SIMULATION ILLUSTRATION

Artificial data allow us to apply the approach in a setting where the DGP is known and so can be judged relative to various assumed theory models. The regressors are stochastic across replications and generated as mutually orthogonal in population by:[3]

$$\mathbf{x}_t \sim \mathsf{IN}_{10}[\mathbf{0}, \mathbf{\Omega}], \tag{28}$$

where $\omega_{ii} = 1$ and $\omega_{ij} = 0, \forall i \neq j$. Two conditional DGPs are considered:

$$\text{Case [A]: } y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \epsilon_t, \tag{29}$$

$$\text{Cases [B] \& [C]: } y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \epsilon_t, \tag{30}$$

where $\epsilon_t \sim \mathsf{IN}[0, 1]$. The coefficients $\beta_i = \psi/\sqrt{T}$ are set to ensure all relevant variables have noncentralities of $\psi = 5, 3, 1$ in the three experiments [A]–[C] described below.

The initial general model is:

$$y_t = \delta_0 + \sum_{j=1}^{10} \delta_j x_{j,t} + \sum_{j=1}^{T} \delta_{j+10} 1_{t=t_j} + \upsilon_t, \tag{31}$$

where $N = 11$, except when the indicators are included for IIS so $N = T + 11$. In case [A], the DGP is (29) where the relevant regressors are $x_{1,t}$ and $x_{2,t}$ so $m = 2$, and these are also the forced regressors in (31), thereby implementing a 'correct' theory as in Section 4.1, while searching over the intercept and $x_{3,t}, \ldots, x_{10,t}$. In case [B], the DGP is now (30), where the relevant regressors are $x_{1,t}, \ldots, x_{4,t}$ so $m = 4$, and the theory again specifies $x_{1,t}$ and $x_{2,t}$ to be forced, which is correct but is incomplete, mimicking Section 4.2. In case [C], the DGP is again (30) but the theory model specifies that $x_{1,t}$ and $x_{5,t}$ are relevant, and hence these are retained, whereas the other relevant variables $x_{2,t}, \ldots, x_{4,t}$ are searched over, so the theory is both incorrect (as $x_{5,t}$ is irrelevant) and incomplete. Selection is undertaken at $\alpha_1 = 0.01$, with diagnostic testing to represent standard usage, with the F-test of the theory model at $\alpha_2 = 0.005$. $T = 100$ and $M = 10,000$ replications are undertaken ($M = 1,000$ for IIS).

Results are calculated excluding any selected indicators when IIS is used. We report the gauge (null retention frequency) for the irrelevant variables retained in

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

the selected model, the potency (retention frequency for relevant variables, where a forced relevant variable has a potency of unity), and the average theory rejection frequency at α_2 (denoted ATRF), defined by:

$$\begin{aligned}
\text{retention rate } \widetilde{\mathsf{p}}_j &= \tfrac{1}{M}\textstyle\sum_{i=1}^{M} 1_{(\widetilde{\delta}_{j,i}\neq 0)}, \quad j=0,\ldots,10;\\
\text{potency} &= \tfrac{1}{m}\textstyle\sum_{j=1}^{m}\widetilde{\mathsf{p}}_j;\\
\text{gauge} &= \tfrac{1}{N-m}\left(\widetilde{\mathsf{p}}_0+\textstyle\sum_{j=m+1}^{N-1}\widetilde{\mathsf{p}}_j\right);\\
\text{ATRF} &= \tfrac{1}{M}\textstyle\sum_{i=1}^{M} 1_{(\mathsf{pF}_i<\alpha_2)},
\end{aligned}$$

where $\widetilde{\delta}_{j,i}$ denotes the OLS coefficient on $x_{j,t}$ in replication i if selected (zero otherwise), and $1_{(\widetilde{\delta}_{j,i}\neq 0)}$ is the indicator variable, equal to unity when the argument is true and zero otherwise. In case [C], however, gauge includes $\widetilde{\mathsf{p}}_5$ only when $|\mathsf{t}_{\delta_5=0}| > 1.5c_{\alpha_1}$, so even though the incorrect theory forces $x_{5,t}$, $m=4$ for calculating potency. Also, pF denotes the p-value of the F-statistic for testing the joint significance of the set of selected, but not forced, regressors in the final model. We also report the percentage of replications in which at least one unforced variable had $|\mathsf{t}_{\delta_j=0}| > 1.5c_{\alpha_1}$, the test in (17) denoted NU, the percentage of replications where no unforced variable was selected (denoted NUVS), and when 2 or more unforced variables were selected (2UVS).

A number of aspects of the analysis in Section 4 are highlighted by the illustration. First, from Table 1, case [A], the theory is of course always retained, even for the noncentrality of unity, a setting where the relevant theory variables would rarely be significant on conventional t-tests even at 5%: see Leeb and Pötscher (2003). The gauge remains close to $\alpha_1 = 0.01$ throughout. Using $|\mathsf{t}_{\delta_i=0}| > 1.5c_{\alpha_1}$, the theory was rejected 0.2% of the time, somewhat larger than derived from (19), as against 3.5% on the simpler F-test. From the analysis in (18), when $\alpha_1 = 0.01$ and $k=9$, then $k\alpha_1 = 0.09$, with $\mathsf{p}_1 \simeq 0.4\%$, which is smaller than 2UVS, whereas $100k\alpha_1 = 9$, so no irrelevant variables will be retained roughly 91% of the time, matching NUVS. In case [B], where four variables are relevant, the costs of not

TABLE 1. Selection with retained theory

	[A]			[B]			[C]		
ψ	5	3	1	5	3	1	5	3	1
Potency	100%	100%	100%	99.1%	80.8%	52.4%	98.6%	71.3%	28.9%
Gauge	1.3%	1.3%	1.3%	1.7%	1.6%	1.3%	1.8%	1.7%	1.3%
NU	0.2%	0.2%	0.2%	95.8%	31.8%	0.6%	98.8%	42.4%	1.0%
ATRF	3.5%	3.5%	3.5%	99.8%	77.8%	8.5%	100%	88.8%	11.5%
NUVS	91.1%	91.1%	91.1%	0.1%	16.2%	84.8%	0.0%	6.4%	80.7%
2UVS	1.7%	1.7%	1.7%	96.7%	44.9%	2.8%	99.7%	68.9%	3.6%

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

TABLE 2. Selection with retained theory using IIS

	[A]			[B]			[C]		
ψ	5	3	1	5	3	1	5	3	1
Potency	100%	100%	100%	98.1%	77.1%	52.5%	97.1%	67.2%	29.1%
Gauge	1.3%	1.3%	1.3%	1.3%	1.6%	1.2%	1.4%	1.5%	1.1%
NU	0.7%	0.7%	0.7%	96.2%	39.8%	1.9%	99.0%	52.1%	2.5%
ATRF	7.8%	7.8%	7.8%	99.6%	73.9%	13.0%	100%	86.4%	15.5%
NUVS	89.3%	89.3%	89.3%	0.3%	23.1%	84.3%	0.0%	10.7%	80.9%
2UVS	0.9%	0.9%	0.9%	93.6%	38.7%	2.2%	99.4%	60.9%	3.3%

forcing retention of two of these are shown by the lower potency, which is halved for $\psi = 1$, and the low rejection of the theory in that case. Conversely, when the theory is seriously wrong ($\psi = 5$), the rejection rates are high. In case [C], so the theory is both incorrect and incomplete, and only one relevant variable is automatically retained, the potency remains high for $\psi = 5, 3$, as does the rejection of the theory, and the gauge remains low.

Despite adding T additional impulse indicators and needing a generalized search, Table 2 shows the results with IIS are close to those without, so the extra costs of search are negligible. Case [A] matches the analysis in Johansen and Nielsen (2009), where the rises in NU and ATRF may be due to not bias correcting the error variance estimator for selected impulse indicators, whereas the drop in 2UVS probably reflects the lower efficiency of the expanding search component of the algorithm.

Overall, while a simple illustration, the findings are consistent with the earlier analysis and suggest that retaining theory while searching over orthogonalized candidate variables and allowing for the possibility of multiple breaks has low costs when the theory is correct and complete, and can reveal defects when it is not. The easily implemented F-test has a reasonable null rejection frequency and good power for both larger noncentralities when the theory is incomplete or incorrect.

6. THE PROBLEMS OF ECONOMIC FORECASTING

The theory of economic forecasting was developed by Haavelmo (1944) for when an econometric model coincided with a stationary DGP. Consider an $n \times 1$ vector $\mathbf{x}_t \sim \mathsf{f}_{\mathsf{x}}(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathsf{R}^k$, where $\mathbf{X}_{t-1} = (\dots \mathbf{x}_1 \dots \mathbf{x}_{t-1})$. A statistical forecast for $\mathbf{x}_{T+h}$ is given by $\widetilde{\mathbf{x}}_{T+h|T} = \mathbf{g}_h(\mathbf{X}_T)$ at a forecast origin T. The key issue concerns how to select $\mathbf{g}_h(\cdot)$, and both his—and the conventional answer—is the conditional expectation:

$$\widehat{\mathbf{x}}_{T+h|T} = E[\mathbf{x}_{T+h} \mid \mathbf{X}_T], \tag{32}$$

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

which is unbiased $E[(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T})|\mathbf{X}_T] = \mathbf{0}$ and has the smallest mean-square forecast-error matrix:

$$\mathsf{M}\left[\widehat{\mathbf{x}}_{T+h|T} \mid \mathbf{X}_T\right] = E\left[\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right)\left(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}\right)' \mid \mathbf{X}_T\right]. \tag{33}$$

So why do forecasts fail so often?

One answer is that econometric models are poor approximations to the conditional expectation. There are nine possible mistakes in a forecasting model, namely: (*i*) equilibrium-mean shifts; (*ii*) slope changes; (*iii*) equilibrium-mean misspecifications; (iv) slope misspecifications; (*v*) forecast-origin mismeasurements; (*vi*) equilibrium-mean misestimation; (*vii*) slope misestimation; (*viii*) omitted variables (including functional form misspecification); and (*ix*) unobserved innovation errors. Although all contribute to worse forecast accuracy than would occur in their absence, in a stationary, ergodic world (so excluding (*i*) and (*ii*)), forecasts from least-squares estimated models unconditionally attain their expected forecast accuracy despite misspecification, as shown by Miller (1978). Consequently, (*iii*)–(*ix*) would not in general induce forecast failure under stationarity. That suggests that nonstationarity is the culprit, and indeed few economic theories allow for unanticipated shifts, although these occur intermittently empirically, as stressed above.

If $\mathbf{x}_t \sim \mathsf{f}_{\mathsf{x}_t}(\mathbf{x}_t|\mathbf{X}_{t-1}, \boldsymbol{\theta})$ changes over $T+1, \ldots, T+h$, then the statement in (32) is not well based and for the conditional expectation to be unbiased, needs to be written as:

$$\widehat{\mathbf{x}}_{T+h|T} = E_{\mathsf{f}_{\mathsf{x}_{T+h}}}[\mathbf{x}_{T+h} \mid \mathbf{X}_T]. \tag{34}$$

That:

$$E_{\mathsf{f}_{\mathsf{x}_T}}[(\mathbf{x}_{T+h} - \widehat{\mathbf{x}}_{T+h|T}) \mid \mathbf{X}_T] = \mathbf{0} \tag{35}$$

is of little relevance when $\mathsf{f}_{\mathsf{x}_{T+h}}(\cdot) \neq \mathsf{f}_{\mathsf{x}_T}(\cdot)$. In particular, as shown in Clements and Hendry (1998) *inter alia*, unanticipated location shifts are pernicious for forecasting as the mean forecast is centered on the incorrect value. While an unanticipatable event obviously cannot be forecast, once a shift has occurred, some models are, and others are not, robust. In particular, EqCMs fail systematically as forecasts continue to revert to the previous (and so incorrect) equilibrium mean: see Castle et al. (2010). However, the difference of an EqCM is robust once past a location shift: see Hendry (2006). Thus, even if location shifts occur just prior to the forecast period, the effort of building, selecting, and evaluating a congruent and encompassing model of an LDGP, retaining the theory-based representation, is not wasted. Within sample breaks can be tackled by IIS, and the difference transform of the selected model can be used to avoid systematic forecast failure, yet retain all the policy implications, prechecked for validity by the appropriate superexogeneity test.

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

7. CONCLUSION

Model selection has had numerous critics from 'data mining' in Lovell (1983) through Leeb and Pötscher (2005). Yet the key implication of the above analysis is that it is almost costless to check large numbers of candidate exogenous variables when retaining a theory-based specification. The retention of the theory variables ensures that there is no selection over the parameters of interest, so that the distributions of their estimates are unaffected by selection over the orthogonalized set of candidates. Under the null that all those candidates are irrelevant, the parameters of interest are unaffected by the reparametrization and therefore by selection.

Conversely, there are substantial benefits if the initial specification is incorrect but the enlarged model nests the data generating process. Our procedure allows an investigator to discover what variables actually matter empirically in addition to those incorporated in the theory, even with endogenous variables and when there are more potential variables than observations, essentially preempting most seminar questions of the form 'did you try my favorite variable(s)?' while controlling the null rejection frequency. We have drawn heavily on Haavelmo's legacy in this variant of empirical model discovery to help achieve his aims.

NOTES

1. Frisch seems to have coined the term 'autonomy' as early as 1931, in a note meant to reach an audience, but which never did (personal communication from Olav Bjerkholt).

2. The forecasting chapter was not in the 1941 version and may have originated when submitting an abstract on prediction for the planned but cancelled December 1942 Econometric Society meeting: see Bjerkholt (2007).

3. We are indebted to Jennifer L. Castle for computing these simulations.

REFERENCES

Aldrich, J. (1989) Autonomy. *Oxford Economic Papers* 41, 15–34.

Aldrich, J. (1994) Haavelmo's identification theory. *Econometric Theory* 10, 198–219.

Anderson, T.W. (1962) The choice of the degree of a polynomial regression as a multiple-decision problem. *The Annals of Mathematical Statistics* 33, 255–265.

Bjerkholt, O. (2005) Frisch's econometric laboratory and the rise of Trygve Haavelmo's probability approach. *Econometric Theory* 21, 491–533.

Bjerkholt, O. (2007) Writing the probability approach with nowhere to go: Haavelmo in the United States, 1939–1944. *Econometric Theory* 23, 775–837.

Bock, M.E., T.A. Yancey, & G.C. Judge (1973) Statistical consequences of preliminary test estimators in regression. *Journal of the American Statistical Association* 68, 109–116.

Castle, J.L., J.A. Doornik, & D.F. Hendry (2011) Evaluating automatic model selection. *Journal of Time Series Econometrics* 3(1); doi:10.2202/1941–1928.1097.

Castle, J.L., J.A. Doornik, & D.F. Hendry (2012) Model selection when there are multiple breaks. *Journal of Econometrics* 169, 239–246.

Castle, J.L., N.W.P. Fawcett, & D.F. Hendry (2010) Forecasting with equilibrium-correction models during structural breaks. *Journal of Econometrics* 158, 25–36.

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

Castle, J.L. & N. Shephard (eds.) (2009) *The Methodology and Practice of Econometrics*. Oxford University Press.

Chatfield, C. (1995) Model uncertainty, data mining and statistical inference (with discussion). *Journal of the Royal Statistical Society, A* 158, 419–466.

Clements, M.P. & D.F. Hendry (1998) *Forecasting Economic Time Series*. Cambridge University Press.

Clements, M.P. & D.F. Hendry (eds.) (2002) *A Companion to Economic Forecasting*. Blackwell.

Clements, M.P. & D.F. Hendry (eds.) (2011) *Oxford Handbook of Economic Forecasting*. Oxford University Press.

Coen, P.G., E.D. Gomme, & M.G. Kendall (1969) Lagged relationships in economic forecasting. *Journal of the Royal Statistical Society A* 132, 133–163.

Davidson, R. & J.G. MacKinnon (2004) *Econometric Theory and Methods*. Oxford University Press.

Dickey, D.A. & W.A. Fuller (1979) Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–431.

Doob, J.L. (1953) *Stochastic Processes*, 1990 ed. John Wiley Classics Library.

Doornik, J.A. (2007) Econometric Model Selection with More Variables than Observations. Working paper, Economics Department, University of Oxford.

Doornik, J.A. (2008) Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics* 70, 915–925.

Doornik, J.A. (2009) Autometrics. See Castle and Shephard (2009), pp. 88–121.

Doornik, J.A. & D.F. Hendry (2009) *Empirical Econometric Modelling using PcGive: Volume I*. Timberlake Consultants Press.

Durbin, J. (1954) Errors in variables. *Review of the Institute of International Statistics* 22, 23–54.

Eddington, C. (1928) *Space, Time, and Gravitation*. Cambridge University Press.

Elliott, G., C.W.J. Granger, & A. Timmermann (eds.) (2006) *Handbook of Econometrics on Forecasting*. Elsevier.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation and testing. *Econometrica* 55, 251–276.

Engle, R.F. & D.F. Hendry (1993) Testing super exogeneity and invariance in regression models. *Journal of Econometrics* 56, 119–139.

Engle, R.F., D.F. Hendry, & J.F. Richard (1983) Exogeneity. *Econometrica* 51, 277–304.

Frisch, R. (1938) Statistical versus theoretical relations in economic macrodynamics. Mimeograph dated 17 July 1938, League of Nations Memorandum. Reproduced by University of Oslo in 1948 with Tinbergen's comments. Contained in Memorandum 'Autonomy of Economic Relations', 6 November 1948, Oslo, Universitets Økonomiske Institutt. Reprinted in D.F. Hendry & M.S. Morgan (1995) *The Foundations of Econometric Analysis*. Cambridge University Press.

Frisch, R. & F.V. Waugh (1933) Partial time regression as compared with individual trends. *Econometrica* 1, 221–223.

Granger, C.W.J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics* 16, 121–130.

Haavelmo, T. (1944) The probability approach in econometrics. *Econometrica* 12, 1–118, supplement.

Haavelmo, T. (1958) The role of the econometrician in the advancement of economic theory. *Econometrica* 26, 351–357.

Haavelmo, T. (1989) *Prize Lecture*. Sveriges Riksbank: Prize in Economic Sciences in Memory of Alfred Nobel.

Hausman, J.A. (1978) Specification tests in econometrics. *Econometrica* 46, 1251–1271.

Hendry, D.F. (2000) *Econometrics: Alchemy or Science?*, New ed. Oxford University Press.

Hendry, D.F. (2006) Robustifying forecasts from equilibrium-correction models. *Journal of Econometrics* 135, 399–426.

Hendry, D.F. (2009) The methodology of empirical econometric modeling: Applied econometrics through the looking-glass. In T.C. Mills & K.D. Patterson (eds.), *Palgrave Handbook of Econometrics*, pp. 3–67. Palgrave MacMillan.

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

Hendry, D.F. (2011a) Empirical economic model discovery and theory evaluation. *Rationality, Markets and Morals* 2, 115–145. Available at http://www.rmm-journal.de/htdocs/st01.html.
Hendry, D.F. (2011b) On adding over-identifying instrumental variables to simultaneous equations. *Economics Letters* 111, 68–70.
Hendry, D.F., S. Johansen, & C. Santos (2008) Automatic selection of indicators in a fully saturated regression. *Computational Statistics* 33, 317–335. Erratum, 337–339.
Hendry, D.F. & H.M. Krolzig (2005) The properties of automatic Gets modelling. *Economic Journal* 115, C32–C61.
Hendry, D.F. & M.S. Morgan (eds.) (1995) *The Foundations of Econometric Analysis*. Cambridge University Press.
Hendry, D.F. & C. Santos (2010) An automatic test of super exogeneity. In M.W. Watson, T. Bollerslev, & J. Russell (eds.), *Volatility and Time Series Econometrics*, pp. 164–193. Oxford University Press.
Hendry, D.F., A. Spanos, & N.R. Ericsson (1989) The contributions to econometrics in Trygve Haavelmo's The probability approach in econometrics. *Sosialøkonomen* 11, 12–17.
Hood, W.C. & T.C. Koopmans (eds.) (1953) *Studies in Econometric Method*. Number 14 in Cowles Commission Monograph. John Wiley & Sons.
Hooker, R.H. (1901) Correlation of the marriage rate with trade. *Journal of the Royal Statistical Society* 64, 485–492.
Hoover, K.D. & S.J. Perez (1999) Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal* 2, 167–191.
Hoover, K.D. & S.J. Perez (2000) Three attitudes towards data mining. *Journal of Economic Methodology* 7, 195–210.
Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.
Johansen, S. (1995) *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
Johansen, S. & B. Nielsen (2009) An analysis of the indicator saturation estimator as a robust regression estimator. See Castle and Shephard (2009), pp. 1–36.
Juselius, K. (1993) VAR modelling and Haavelmo's probability approach to econometrics. *Empirical Economics* 18, 595–622.
Keynes, J.M. (1939) Professor Tinbergen's method. *The Economic Journal* 44, 558–568.
Keynes, J.M. (1940) Statistical business-cycle research: Comment. *The Economic Journal* 50, 154–156.
Koopmans, T.C. (1947) Measurement without theory. *Review of Economics and Statistics* 29, 161–179.
Koopmans, T.C. (ed.) (1950) *Statistical Inference in Dynamic Economic Models*. Number 10 in Cowles Commission Monograph. John Wiley & Sons.
Krüger, L., G. Gigerenzer, & M.S. Morgan (eds.) (1987) *The Probabilistic Revolution*, vol. 2. MIT Press.
Kurcewicz, M. & J. Mycielski (2003) A Specification Search Algorithm for Cointegrated Systems. Discussion paper, Statistics Department, Warsaw University.
Leamer, E.E. (1978) *Specification Searches. Ad-Hoc Inference with Non-Experimental Data*. John Wiley.
Leeb, H. & B.M. Pötscher (2003) The finite-sample distribution of post-model-selection estimators, and uniform versus non-uniform approximations. *Econometric Theory* 19, 100–142.
Leeb, H. & B.M. Pötscher (2005) Model selection and inference: Facts and fiction. *Econometric Theory* 21, 21–59.
Liao, Z. & P.C.B. Phillips (2012) Automated Estimation of Vector Error Correction Models. Discussion paper, Economics Department, Yale University.
Lovell, M.C. (1983) Data mining. *Review of Economics and Statistics* 65, 1–12.
Lucas, R.E. (1976) Econometric policy evaluation: A critique. In K. Brunner & A. Meltzer (eds.), *The Phillips Curve and Labor Markets. Carnegie-Rochester Conferences on Public Policy*, vol. 1, pp. 19–46. North-Holland Publishing Company.

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press

Marschak, J. & O. Lange (1940) Mr. Keynes on the statistical verification of business cycle theories. See Hendry and Morgan (1995).
Miller, P.J. (1978) Forecasting with econometric methods: A comment. *Journal of Business* 51, 579–586.
Mills, T.C. (2010) Bradford Smith: An econometrician decades ahead of his time. *Oxford Bulletin of Economics and Statistics* 73, 276–285.
Moene, K.A. & A. Rødseth (1991) Nobel laureate: Trygve Haavelmo. *Journal of Economic Perspectives* 5, 175–192.
Morgan, M.S. (1990) *The History of Econometric Ideas*. Cambridge University Press.
Omtzig, P. (2002) Automatic identification and restriction of the cointegration space. Thesis chapter, Economics Department, Copenhagen University.
Pagan, A.R. (1987) Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys* 1, 3–24.
Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.
Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59, 283–306.
Popper, K.R. (1959) *The Logic of Scientific Discovery*. Basic Books.
Popper, K.R. (1963) *Conjectures and Refutations*. Basic Books.
Rapach, D.E. & M.E. Wohar (eds.) (2008) *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Emerald Group.
Sargan, J.D. (2001) Model building and data mining. *Econometric Reviews* 20, 159–170. First presented to the Association of University Teachers of Economics, Manchester, 1973.
Sims, C.A., J.H. Stock, & M.W. Watson (1990) Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.
Smith, B.B. (1926) Combining the advantages of first-difference and deviation-from-trend methods of correlating time series. *Journal of the American Statistical Association* 21, 55–59.
Smith, B.B. (1927) Forecasting the volume and value of the cotton crop. *Journal of the American Statistical Association* 22, 442–459.
Smith, B.B. (1929) Judging the forecast for 1929. *Journal of the American Statistical Association* 24, 94–98.
Spanos, A. (1989) On re-reading Haavelmo: A retrospective view of econometric modeling. *Econometric Theory* 5, 405–429.
Tinbergen, J. (1939) *Statistical Testing of Business-Cycle Theories, vol. I, A Method and Its Application to Investment Activity*. League of Nations.
Toda, H.Y. & P.C.B. Phillips (1993) Vector autoregressions and causality. *Econometrica* 61, 1367–1393.
Vining, R. (1949) A rejoinder. *Review of Economics and Statistics* 31, 91–94.
Wu, D. (1973) Alternative tests of independence between stochastic regressors and disturbances. *Econometrica* 41, 733–750.
Yule, G.U. (1926) Why do we sometimes get nonsense-correlations between time-series? A study in sampling and the nature of time series (with discussion). *Journal of the Royal Statistical Society* 89, 1–64.

https://doi.org/10.1017/S0266466614000218 Published online by Cambridge University Press